APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 1990, p. 3117–3124
0099-2240/90/103117-08$02.00/0

Vol. 56, No. 10

celB, a Gene Coding for a Bifunctional Cellulase from the Extreme Thermophile "*Caldocellum saccharolyticum*"

D. J. SAUL,* L. C. WILLIAMS, R. A. GRAYLING, L. W. CHAMLEY, D. R. LOVE, AND P. L. BERGQUIST

Centre for Gene Technology, Department of Cellular and Molecular Biology, University of Auckland, Auckland, New Zealand

Received 18 April 1990/Accepted 8 August 1990

"*Caldocellum saccharolyticum*" is an obligatory anaerobic thermophilic bacterium. A gene from this organism, designated *celB*, has been cloned in *Escherichia coli* as part of a bacteriophage λ gene library. This gene produces a thermostable cellulase that shows both endoglucanase and exoglucanase activities on test substrates and is able to degrade crystalline cellulose to glucose. The sequence of *celB* has homology with both exo- and endoglucanases described by others. It appears to have a central domain without enzymatic activity which is joined to the enzymatic domains by runs of amino acids rich in proline and threonine (PT boxes). Deletion analysis shows that the exoglucanase activity is located in the amino-terminal domain of the enzyme and that endoglucanase activity is located in the carboxy-terminal domain. There are internal transcriptional and translational start sites within the gene. The intact gene has been cloned into a temperature-inducible expression vector, pJLA602, and overexpressed in *E. coli*. Polyacrylamide gel electrophoresis showed that *celB* produced a protein with a molecular weight of 118,000 to 120,000. A number of smaller proteins with activity against carboxymethyl cellulose and 4-methyl umbelliferyl-β-D-cellobioside were also produced. These are believed to be the result of alternative translational start sites and/or proteolytic degradation products of the translated gene product.

The cellulase system comprises three general classes of enzymes: exoglucanases (β-1,4-D-glucan cellobiohydrolase), endoglucanases (β-1,4-D-glucan glucanhydrolase), and β-1,4-D-glucosidase. The first two enzymes depolymerize cellulose to cellobiose and oligosaccharides, and β-glucosidase then hydrolyzes these sugars to form glucose (1, 31). Most studies of cellulase production have centered on the fungus *Trichoderma reesei* (1, 31), and several components of the cellulase system from this organism have been cloned (4, 20, 40; for a review, see reference 8). Recently, interest has shifted to cellulolytic bacteria. The thermophilic bacterium *Clostridium thermocellum* produces highly active and relatively thermostable enzymes, and several groups have isolated endoglucanase genes from genomic libraries of this organism (1, 9, 10, 39). Béguin and his collaborators have sequenced three endoglucanases from *Clostridium thermocellum* and found that each of the *celA*, *-B*, and *-D* genes have a potential signal peptide sequence resembling signal sequences in gram-positive bacteria, although the peptides show little similarity to each other. Each of the enzymes shares a conserved, reiterated domain coding for the carboxy terminus of the gene and have recognizable promoter-like sequences and ribosomal binding sites (3, 15, 18). Similar results have been reported for *celE* (17).

The genes *cex* and *cenA* from *Cellulomonas fimi* have been cloned and sequenced (33, 46). The gene *cex* encodes for an exoglucanase, and *cenA* codes for an endoglucanase. Each enzyme was shown to be composed of three regions: a short sequence of Pro-Thr residues, a region of hydroxyamino acids that have a low-charge density and are about 50% conserved, and a poorly conserved region which includes most of the polypeptide and contains the putative active site. Warren et al. (44) have proposed that the nonconserved regions represent the catalytic sites and have shown that when the genes *cenA* and *cex* are fused genetically, a hybrid protein is produced which displays the catalytic properties of both enzymes.

"*Caldocellum saccharolyticum*" is an extremely thermophilic obligate anaerobe which will grow and show cellulolytic activity at 80°C. It is able to degrade cellulose but showed no homology to *Clostridium thermocellum* in DNA-DNA hybridization studies (12). We have constructed a gene bank of the chromosomal DNA of this organism and have isolated λ recombinants in which the genes for cellulose breakdown are expressed in *E. coli*. We have described previously the isolation and expression of genes coding for β-glucosidase and a xylanase from "*Caldocellum saccharolyticum*" (25; 27). These enzymes have remarkable temperature stability; temperature optima for activity are between 70 and 85°C. In this communication we describe the characterization and features of the sequence of a bifunctional protein that has both endo-β-1,4-glucanase and exo-β-1,4-glucanase activities and appears to be a naturally occurring homolog of the gene fusion constructed by Warren et al. (44).

MATERIALS AND METHODS

Bacterial strains. PB1427 is strain C600 *thi-1 thrA1 leuB6 lacY1 supE44 tonA* F^-. PB2946 is strain JM83 *ara* Δ(*lac pro*) *thi strA lacZ*ΔM15 *phi*80Δlac F^-. CAG629 is *lon htpR165*-Tn*10* (Carol Gross, Department of Bacteriology, University of Wisconsin, personal communication).

Genomic library. Chromosomal DNA from "*Caldocellum saccharolyticum*" was isolated and partially digested with *Sau*3A. Fragments (10 to 14 kilobases [kb]) were purified and ligated into *Bam*HI-cut λ 1059 (19) or λ 2001 (32) and packaged in vitro. Recombinant DNA techniques were as previously described (25, 32) with pUC and pBS vectors (41). Bacterial strains, media, and the preparation of the size-fractionated thermophilic DNA have been described previously (25).

Restriction endonuclease mapping of λ isolates. *Bam*HI and

* Corresponding author.

*Hin*dIII restriction maps were made by using the *cos* end-labeling method of Rackwitz et al. (34).

Screening for cellulase-positive cloned fragments. Plaques or colonies were plated on defined media and incubated at 37°C for 12 h. The plates were overlaid with 0.7% agarose containing 0.5% carboxymethyl cellulose (CMC) and were shifted to 70°C for 4 to 12 hours. The *E. coli* cells lysed and the active enzyme diffused into the substrate layer in the plate. CM-cellulase (CMCase) activity was recognized by zones of clearing after the plates were stained with Congo Red and destained with 1 M NaCl (43). Exoglucanase activity was assayed by overlaying colonies or plaques with 0.4% soft agar containing 4-methyl umbelliferyl-β-D-cellobioside (MUC; 30, 43) and observing those plaques or colonies which fluoresced at 340 nm after 6 h of incubation at 70°C. Because of the low specificity of this substrate, we have referred to activity on it as 4-methyl umbelliferyl-β-D-cellobiosidase (MUCase).

Dideoxy sequencing. The method of Sanger et al. (35) was used with the single-stranded bacteriophage vectors mp18 and mp19 and the universal sequencing primer. Double-stranded sequencing (5) failed to produce reliable results for large portions of the thermophilic DNA insert, and so where possible, deletions in pBS(−) were converted to single-stranded DNA by infection of the cells with helper phage. A further six oligonucleotide primers were synthesized for confirmation of the sequence on both strands.

Enzyme assays. The release of reducing sugars (glucose equivalents) from the enzymatic hydrolysis of CMC was determined by the method of Lever (23). Protein concentrations were determined according to the method of Lowry et al. (26). Exoglucanase activity was determined on partially purified samples with Avicel as a substrate, and measurement of reducing sugar release was as described above. Fluorometric assays of activity on MUC were carried out by the method of Chernoglaznov et al. (6).

Visualization of enzyme activities on gels. The method of Béguin (2) was modified as followed. Cellular protein (150 μg) was separated on a 7.5% Laemmli gel (21) with 0.2% CMC included in the separating layer. Sodium dodecyl sulfate was removed by washing the gel at room temperature in two changes of a solution containing 10 mM Tris (pH 7.5), 5 mM β-mercaptoethanol, and 20% isopropanol. Renaturation of proteins was carried out by rocking the gel overnight in a solution containing 50 mM Tris (pH 7.5), 5 mM β-mercaptoethanol, and 1 mM EDTA at 4°C. Subsequently, gels were soaked in 50 mM sodium phosphate buffer (pH 6.8) for 1 h, transferred to fresh buffer containing 100 μg of MUC per ml, and soaked for a further 2 h at 4°C. Gels were then transferred onto a glass plate, sealed in film, and incubated at 70°C for 30 min. MUCase activity was visualized as fluorescent bands under UV light. Gels were incubated for a further 4 h, stained in 1% Congo Red–5 mM NaOH for 1 h, and destained overnight in 1 M NaCl–5 mM NaOH. Clear bands indicate the presence of CMCase activity.

Insertion of the 4.85-kb CMC^+ MUC^+ fragment into pBR322: λ NZP2 was derived from a plaque that clearly expressed CMCase and MUCase activity when assayed at 70°C on plates containing CMC and MUC (see Results). DNA from this phage was mapped for restriction sites. A *Bam*HI site has been fortuitously restored at the left-hand end of the inserted DNA (Fig. 1A). λ NZP2 was digested with *Bam*HI, the fragments were ligated into *Bam*HI-digested pBR322, and the resulting mixture was used to transform strain PB1427 to ampicillin resistance. A 4.85-kb fragment of "*Caldocellum saccharolyticum*" DNA was capable of expressing both CMCase and MUCase activities with the insert present in either orientation in the vector (pNZ1026 or pNZ1027).

Cloning of the 4.85-kb CMC^+ MUC^+ fragment in pBS and deletion analysis. The 4.85-kb *Bam*HI "*Caldocellum saccharolyticum*" insert of λ NZP2 was excised and ligated into pBS(−) (41) to create pNZ1087 and into pBS(+) to create pNZ1088. This insert contains the gene designated *celB*. The orientation was determined with reference to the single *Pst*I sites in the insert and the vector. The "*Caldocellum saccharolyticum*" DNA present in pNZ1087 DNA was deleted in either a rightwards or leftwards direction after appropriate digestion of purified DNA with exonuclease III and mung bean nuclease, as described previously (36). Other deletions were made by using suitable restriction enzyme cleavage sites.

Construction of recombinant plasmids with the expression vector pJLA602. The plasmid pJLA602 has the λ p_L promoter controlled by the λ *c*I857 temperature-sensitive repressor. Induction of the promoter is achieved by raising the growth temperature from 32 to 42°C (38). Three *celB* recombinants were made by using this vector in strain PB1427. (i) pNZ1100 is the entire 4.85-kb *celB* *Bam*HI-*Bam*HI fragment of λ NZP2 inserted into the *Bam*HI site of the vector. This construction contains the complete open reading frame with 1.2 kilobase pairs of intervening sequence between the p_L promoter of the vector and the start codon of *celB*. (ii) pNZ1101 is an *Nde*I deletion of pNZ1100. This construction deletes 0.9 kilobase pairs of the intervening sequence between the p_L promoter of the vector and the open reading frame. (iii) pNZ1102 is a *Nco*I deletion of pNZ1100. This construction deletes all the intervening sequence and the coding region for domain A of the putative peptide. This plasmid was made to locate any alternative start sites within the open reading frame.

Nucleotide sequence accession number. The complete sequence of *celB* has been determined (37) and is available from the EMBL Data Library under accession no. X13602.

RESULTS

λ Recombinants with CMCase and MUCase activity. λ libraries of "*Caldocellum saccharolyticum*" were screened for activities on CMC and MUC substrates. Three phenotypes were observed: CMC^+ MUC^-, CMC^- MUC^+, and CMC^+ MUC^+. CMCase activity was assumed to signify that the recombinants coded for an endoglucanase. MUCase activity could have occurred as the result of a number of different enzyme activities. For example, some xylanases cleave MUC but have no activity on crystalline cellulose (27). Biochemical analysis of partially purified enzymes from some of the plasmids constructed from λ NZP2 showed that the $CMCase^+$ phenotype was consistent with β-1,4-D-glucan glucanhydrolase activity and that the $MUCase^+$ phenotype represented a β-1,4-D glucan cellobiohydrolase activity as determined by using crystalline cellulose (Avicel) as a substrate (Lynly Schofield and R. M. Daniel, University of Waikato, personal communication). We have used the names endoglucanase and exoglucanase in the text to describe these activities.

Restriction enzyme mapping and hybridization studies showed that some λ recombinants overlapped and that more than one gene coding for endoglucanase was present in "*Caldocellum saccharolyticum*," some of which were closely linked and had nearly identical restriction enzymes digestion patterns although they were sited in different

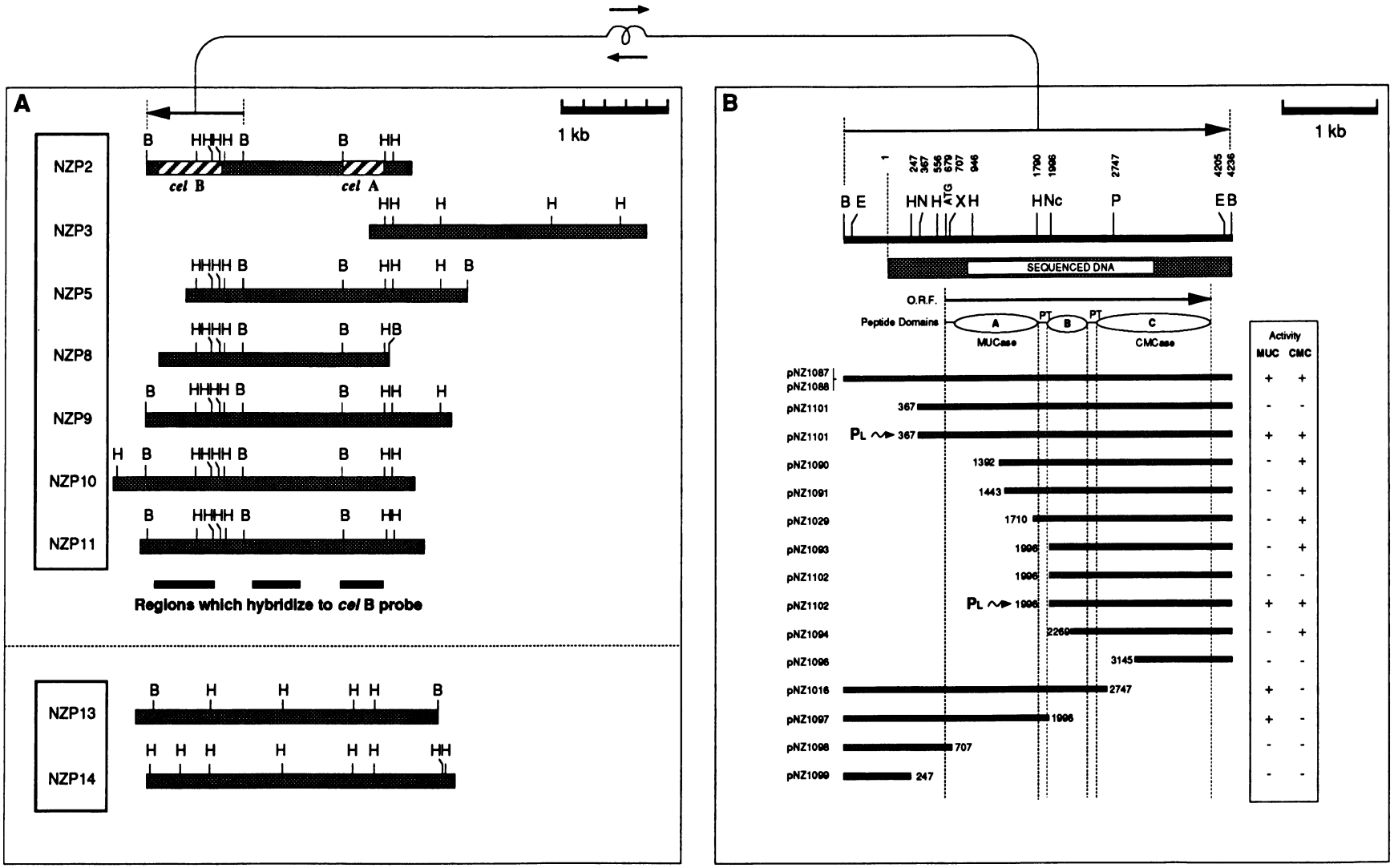

FIG. 1. (A) Restriction maps of λ recombinants. The location of the *Bam*HI and *Hin*dIII sites in the inserts of nine isolates from a λ genomic library of "*Caldocellum saccharolyticum*." The *celB* and *celA* open reading frames (▨) on NZP2 are highlighted. Regions of isolates NZP2 through to NZP11 which hybridize to a *celB* probe are shown as black bars. The 4.85-kb *Bam*HI-*Bam*HI fragment of NZP2 which was used in subsequent steps is also indicated. NZP13 and NZP14 hybridized weakly to a *celB* probe but did not show CMCase and/or MUCase activity. (B) Diagrammatic representation of deletion mutations in the *celB* gene and their enzymatic activities. CMCase and MUCase activities were measured by plate assays, as described in Materials and Methods. The numbers refer to the position in the sequence (37) at which the deletions commence. A diagrammatic representation of the peptide domains and their activities is included. Plasmids pNZ1087 to pNZ1099 were constructed in pBS; pNZ1016 and pNZ1029 were constructed in pBR322. pNZ1101 and pNZ1102 are recombinants of *celB* in the expression vector pJLA602. These are displayed with the p_L promoter uninduced and induced (⇝). B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; N, *Nde*I; Nc, *Nco*I; P, *Pst*I; X, *Xho*I.

contexts within the genome (Fig. 1A). Preliminary estimates made by using the *celB* sequence as a probe suggested that there were five genes with significant similarity in a random library in which any gene was present at a probability of 99.9% (8). Of 1,500 λ recombinants screened in this library, 6 λ recombinants were scored as CMC^+ MUC^+ and 3 were scored as CMC^+ MUC^-.

Enzyme activities of *celB* deletion mutants. A series of subclones and deletion derivatives of *celB* were tested for their CMCase and MUCase activities (Fig. 1B). The results show that the two activities were independent and map the exoglucanase activity to domain A and the endoglucanase to domain C. These results are in accord with expectations from homology comparisons (discussed below). CMCase activity was expressed by several plasmids which have lost the ability to hydrolyze MUC; for pNZ1029, which has been investigated in detail, the temperature optimum (85°C) and stability of the enzyme (half-life, 29 h at 70°C) were unchanged from those produced by the undeleted plasmid pNZ1088. Conversely, deletions of the 3′ portion of *celB* retained MUCase activity and lost CMCase activity. The enzyme produced by pNZ1016, similar to that of pNZ1029, maintains the stability and temperature optimum of the complete enzyme (Lynly Schofield and R. M. Daniel, personal communication). The purpose of domain B is less clear than those of A and C. Deletions into this domain did not affect MUCase or CMCase activities. However, pNZ1016 (domains A and B present) had activity on Avicel, whereas pNZ1097 (domain A) had none (data not shown). Quantitative results are unreliable because there were variable forms of the gene product present in the assay mixture (discussed below), but this result suggests that the central domain may be required for activity on insoluble cellulose.

Plasmids pNZ1100 and pNZ1101 were constructed in the expression vector pJLA602 (Materials and Methods). Unlike pNZ1100, pNZ1101 was incapable of expression in the uninduced state. This result indicates that the promoter used by *E. coli* lies within the thermophilic insert upstream of the *Nde*I site at position 367. However, pBS deletions beyond that point maintained expression, suggesting that an alternative promoter is provided by the pBS vector. There are eight deletions in pBS which extended into the open reading frame but still expressed CMCase activity. We conclude from inspection of the sequences of the vector and the deletions that these are not in-frame fusions with the vector, and hence a secondary ribosome binding site and start site must exist between positions 2269 and 3145 (the deletion start points of pNZ1094 and pNZ1096, respectively).

The plasmid pNZ1102 gave an anomalous result on MUC when in the induced state. This observation is discussed below.

Analysis of the DNA sequence. The 4,241-base-pair sequence contains one large open reading frame of 3,117 base pairs which is responsible for coding both enzymatic activities. The overall GC content is 37%, which is similar to that found for "*Caldocellum saccharolyticum*" (34%) (12) and also for the other sequences published for this organism: the β-glucosidase gene (38% GC) (24) and a fragment containing genes involved in hemicellulose degradation (35% GC) (27). The GC content increases within the large open reading frame, presumably resulting from constraints enforced in coding regions. Flanking regions, however, are highly AT rich with localized areas in excess of 75% A or T. The sequence upstream of the large open reading frame is distinctive for its long stretches of A's or T's (37). A Shine-Dalgarno sequence is appropriately spaced upstream of the open reading frame.

Analysis of the deduced CelB peptide sequence. The deduced CelB peptide is 1,039 amino acids long and has a calculated nonglycosylated molecular mass of approximately 118 kilodaltons. The region between amino acids 300 and 600 consists almost entirely of β-sheets, as predicted by computer analysis (7, 13). A similar observation has been made in *Trichoderma reesei* (20), although no clear function of this structure in cellulases is as yet known.

There are two distinctive domains on the peptide sequence with successive Pro-Thr repeats, which have been observed to occur in several other cellulases. These are designated PT boxes by Langsford et al. (22). The first starts at residue 379 and consists of 43 amino acids, and the second, at residue 583, is 45 amino acids long. These PT boxes effectively divide the protein into three major domains designated A, B, and C.

At the amino terminus of the putative peptide is a short hydrophilic region with charged residues followed by a longer hydrophobic domain. This sequence is very similar to leader sequences found in other gram-positive bacteria (28). At the carboxy terminus of this hydrophobic region is a sequence which conforms to the rules governing leader sequence cleavage recognition sites (11).

Similarities between the putative amino acid sequence of "*Caldocellum saccharolyticum*" CelB and other known cellulases are shown in Fig. 2.

Detection of *celB* gene product. (i) Overexpression of *celB*. The three plasmids pNZ1100, pNZ1101, and pNZ1102 are described in Materials and Methods. Table 1 shows the MUCase activities of strain PB1427 carrying these plasmids and grown at 32 and 42°C. These activities are calculated with respect to the uninduced value for pNZ1100, that is, as a multiple of the expression from the native promoter. Activities of induced pNZ1100 and pNZ1101 were similar; both gave an approximately 40-fold increase in expression as compared with pNZ1100 at 32°C. pNZ1102 also had considerable activity on MUC despite the fact that this plasmid had been deleted for domain A. This activity was most likely the result of activity from domain C. This observation emphasizes that MUC is hydrolyzed to various degrees by a broad range of enzymes and is not a specific substrate for exoglucanase. This activity was not detected by plate assays with pBS constructs, but the high level of expression in pJLA602-derived plasmids was sufficient to give positive MUCase results when tested in this way. This accounts for the MUC^+ result of induced pNZ1102 (Fig. 1B).

(ii) Activity gels. To visualize the *celB* gene product in a background of total cellular protein, MUC and CMC substrates were incorporated into the polyacrylamide gels and activities of bands were detected directly (see Materials and Methods). Figure 3 shows the profiles of total cellular protein of strain PB1427 transformed with the pJLA602 vector pNZ1100 or pNZ1101. MUCase and CMCase activities were tested in situ for the three strains.

The highest molecular mass band produced by pNZ1100 on the activity gels corresponds to 118 to 120 kilodaltons. This band is also visible on the Coomassie blue-stained gel (Fig. 3, arrows), and its size is in good agreement with the size of the protein calculated from sequence data. The banding pattern for CMC and MUC was similar for the high-molecular-weight bands. This reflects the abilities of both domains A and C to hydrolyze MUC. Differences only occurred among proteins of lower molecular mass. Bands in the pNZ1102 lane confirm that initiation of translation occurred with the 5′ end of the open reading frame absent. The

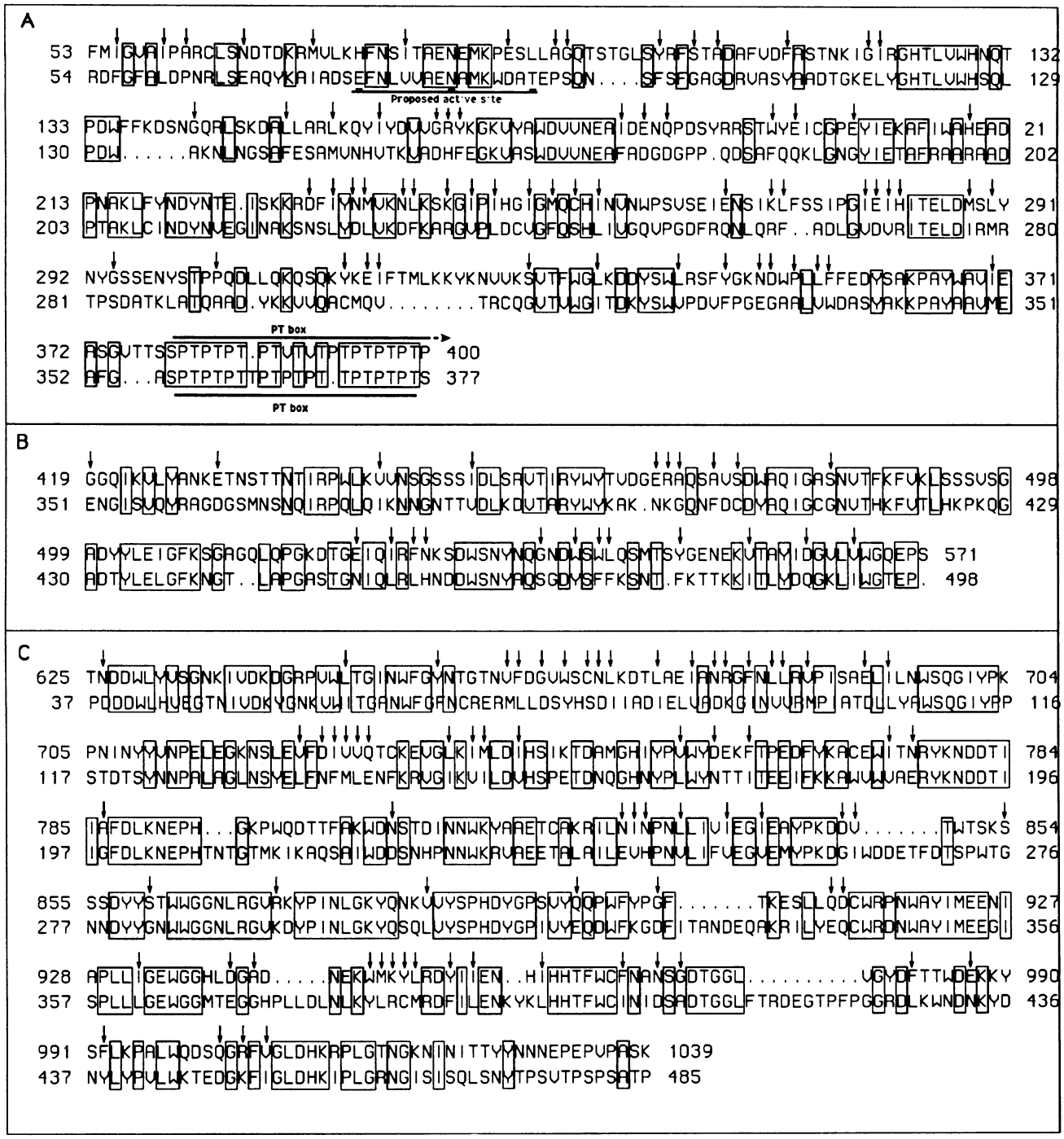

FIG. 2. Comparison of "*Caldocellum saccharolyticum*" CelB protein with other published cellulases. Identical amino acids are boxed. Conservative changes are labeled with arrows. The top strand in each panel is the deduced peptide sequence from "*Caldocellum saccharolyticum*." (A) Comparison with Cex peptide of *Cellulomonas fimi*. Also indicated are the PT boxes and the proposed active site of the *Cellulomonas* protein (28). (B) Comparison with the endoglucanase of *Bacillus subtilis* (27). (C) Comparison with EGB peptide of *Clostridium thermocellum* (15).

highest-molecular-weight band for this plasmid was approximately 82 kilodaltons. This result is consistent with a translational start at a GTG codon at position 2011.

It is as yet uncertain whether the large number of bands was caused by multiple sites for the initiation of translation or by proteolytic degradation. The banding pattern observed was not reduced by phenlymethylsulfonyl fluoride or the use of CAG627, a deficient strain for heat shock proteins. If proteolysis is involved, then the molecular masses of the fragments are not explained by the proposition that the major sites of attack are in the regions of the PT boxes, as suggested by others (22).

TABLE 1. MUCase activities of strains carrying expression vector constructions containing the *celB* gene[a]

Plasmid	MUCase activity at:	
	32°C	42°C
pNZ1100	1.00	36.0
pNZ1101	0.13	27.8
pNZ1102	0.05	14.4

[a] Plasmids were tested in the induced and uninduced state and were assayed as previously described (6). Values were calculated as a fraction of the *celB* gene activity controlled by its native promoter (pNZ1100 at 32°C). All values are corrected for a background activity of the strain carrying the vector alone. This value becomes significant when MUC is incubated for prolonged periods at 70°C. Under the conditions described, background activity accounted for 10.5% of the activity value measured for native promoter expression.

DISCUSSION

The *celB* gene is novel because of its bifunctional nature, and it appears to be a natural homolog of the construction engineered by Warren and his collaborators (44). Domain A of CelB has 34% homology with the cellobiohydrolase Cex sequence from *Cellulomonas fimi* (33). This homology is up to and including the PT boxes of both proteins. The homology gives both proteins a similar hydropathic profile, implying a similar overall structure.

The Cex peptide and the CenA peptide of *Cellulomonas fimi* are split by a PT box into two domains. The C-terminal domain of Cex has homology with the N-terminal domain of the CenA protein. The putative protein described here had homology only with the Cex N-terminal domain. This is reflected by a lack of detectable similarity between CelB and CenA.

Miller et al. (29) have suggested a possible active site for the Cex protein on the basis of the active site of hen egg white lysozyme. Both Cex and CenA have the required Glu-7x-Asn-6x-Thr consensus (x meaning any amino acid). Despite the high homology between Cex and CelB in this region, two of the three essential amino acids are absent at the corresponding point of CelB, and this consensus cannot be found elsewhere in the protein. West et al. (45) have also pointed out that *Bacillus* and *Trichoderma* cellulases do not contain the Glu-7x-Asn-6x-Thr sequences, and they believe that this identification of the catalytic site of *Cellulomonas fimi* is incorrect. Homology of this region with hemicellulase-degrading enzymes is described by Lüthi et al. (27).

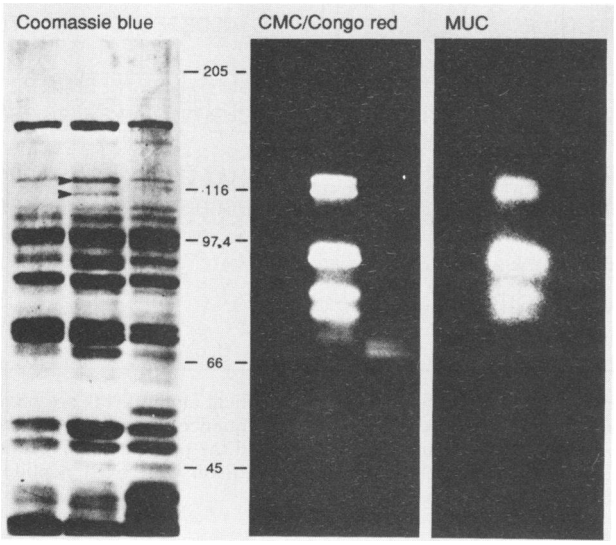

FIG. 3. Protein profiles and activity gels of *E. coli* PB1427 carrying different *celB* recombinant plasmids. Left lanes, pJLA602 vector; center lanes, pNZ1100; right lanes, pNZ1102. The CMC-Congo Red and MUC activity gels were prepared as described in the text. The two largest protein bands which are unique to pNZ1100 and which are visible in the Coomassie-stained gel are labeled with arrows. Molecular mass markers are shown in kilodaltons.

There is 45% homology between domain B of CelB and the *Bacillus subtilis* endoglucanase described by MacKay et al. (28). This portion of the *B. subtilis* sequence is outside the short region in which MacKay et al. noted similarities with part of the *Clostridium thermocellum celB* endoglucanase gene and also outside the area noted as having homology with a cellulase from *Erwinia chrysanthemi* and alkalophilic *Bacillus* sp. strain N = 4 (16). The homology is with the C-terminal domain of the *Bacillus* enzyme and begins close to the short PT box of this protein. A 51% homology can be found between the C-terminal domain of "*Caldocellum saccharolyticum*" CelB and the EGB endoglucanase of *Clostridium thermocellum* (the gene of which is also designated *celB*) (16). This region includes the short span of *B. subtilis-Clostridium thermocellum* homology described above but excludes the C-terminal reiterated region present in both EGA and EGB of *Clostridium thermocellum* (Fig. 2).

In summary, our results show that the individual domains were active on artificial substrates and that there was significant homology with the catalytic domains of other cellulase sequences. However, there was little sequence similarity with the binding region of Cex and no homology with CenA. The central domain had some homology with an endoglucanase from *B. subtilis*, and the third domain showed homology with the CelB endoglucanase of *Clostridium thermocellum*. The amino-terminal portion of the gene also showed homology with a bifunctional xylanase-β-xylosidase isolated from "*Caldocellum saccharolyticum*" (27). It appears likely that both xylanase and cellulase activities have evolved from a common ancestral gene, a conclusion that is supported by amino acid homology comparisons of West et al. (45) with other cellulases and xylanases.

Warren and his collaborators have provided evidence that the PT boxes are sites of glycosylation for CenA and Cex (14, 22). Degradation of CelB by proteases may be occurring in *E. coli*, since in this host it would not be glycosylated. Work is in progress to express this gene in *Saccharomyces cerevisiae* and to determine whether a more stable, glycosylated product can by synthesized. We have not been able to examine directly the *celB* gene product from "*Caldocellum saccharolyticum*" because of difficulties in isolating it from the multiplicity of cellulases present in extracts of this organism (R. M. Daniel, personal communication).

Several cellulase genes have been isolated from "*Caldocellum saccharolyticum*," and some λ CMC^+ recombinants from the gene library do not hybridize to a *celB* probe (R. Grayling, unpublished M.S. thesis, University of Auckland, Auckland, New Zealand, 1990). Clearly "*Caldocellum saccharolyticum*," like other cellulolytic organisms, utilizes a number of enzyme families to degrade native cellulose. It is of interest that the deletion mutants carrying one of the enzymatic domains retain activity. This suggests that *celB* may have resulted from a gene fusion, and this may explain why we have been unable to isolate an exocellulase from our genomic library of "*Caldocellum saccharolyticum*."

ACKNOWLEDGMENTS

This work was supported in part by grants from the University of Auckland Research Committee and from Pacific Enzymes Ltd.

LITERATURE CITED

1. **Aubert, J. P., P. Béguin, and J. Miller (ed.).** 1988. Biochemistry and genetics of cellulase degradation. FEMS symposium no. 43. Academic Press, New York.
2. **Béguin, P.** 1983. Detection of cellulase activity in polyacrylamide gels using Congo red-stained agar replicas. Anal. Biochem. **131**:333–336.
3. **Béguin, P., P. Cornet, and J.-P. Aubert.** 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. J. Bacteriol. **162**:102–105.
4. **Bhikhabhai, R., G. Johansson, and L. G. Pettersson.** 1984. Isolation of cellulolytic enzymes from *Trichoderma reesei* QM9414. J. Appl. Biochem. **6**:336–345.
5. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for screening plasmid DNA. DNA **4**:165–170.
6. **Chernoglaznov, V. M., A. N. Jatarova, and A. A. Klyosov.** 1989. Continuous photometric determination of endo-1,4-β-D-glucanase (cellulase) activity using 4-methylumbelliferyl-β-D-cellobioside as substrate. Anal. Biochem. **179**:186–189.
7. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequences. Adv. Enzymol. Relat. Areas Mol. Biol. **47**:45–148.
8. **Clark, L., and J. Carbon.** 1976. A colony bank containing synthetic ColEI hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–100.
9. **Cornet, P., D. Tronik, J. Millet, and J.-P. Aubert.** 1983. Cloning and expression in *E. coli* of *Clostridium thermocellum* genes coding for amino acid synthesis and cellulose hydrolysis. FEMS Microbiol. Lett. **16**:137–141.
10. **Coughlan, M. P. (ed.).** 1989. Enzyme systems for lignocellulose degradation. Commission of the European Community, Elsevier Applied Science Publishers Ltd., London.
11. **Dierstein, R., and W. Wickner.** 1986. Requirements for substrate recognition by bacterial leader peptidase. EMBO J. **5**:427–431.
12. **Donnison, A. M., C. M. Brocklesbury, and W. H. Morgan.** 1986. A new species of non-sporulating thermophilic cellulolytic bacterium. Proc. Int. Congr. Microbiol. (Manchester) 203.
13. **Garnier, J.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
14. **Gilkes, N. R., R. A. J. Warren, R. C. Miller, and D. G. Kilburn.** 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. **263**:10401–10407.
15. **Grepinet, O., and P. Béguin.** 1986. Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B. Nucleic Acids Res. **14**:1791–1799.
16. **Guiseppi, A., B. Cami, J.-L. Aymeric, G. Ball, and N. Creuzet.** 1988. Homology between endoglucanase Z of *Erwinia chrysanthemi* and endoglucanases of *Bacillus subtilis* and alkalophilic *Bacillus*. Mol. Microbiol. **2**:159–164.
17. **Hall, J., G. P. Hazlewood, P. J. Baker, and H. G. Gilbert.** 1988. Conserved reiterated domains of *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. Gene **69**:30–38.
18. **Joliff, G., P. Béguin, and J.-P. Aubert.** 1986. Nucleotide sequence of the cellulase gene *celD* encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. **14**:8605–8613.
19. **Karn, J., S. Brenner, L. Barnett, and G. Cesarini.** 1980. Novel bacteriophage λ cloning vector. Proc. Natl. Acad. Sci. USA **77**:5172–5176.
20. **Knowles, J. K. C., P. Lehtovaara, and T. T. Teeri.** 1987. Cellulase families and their genes. Trends Biotechnol. **5**:255–260.
21. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
22. **Langsford, M. L., N. R. Gilkes, B. Singh, B. Morer, R. C. Miller, R. A. J. Warren, and D. G. Kilburn.** 1987. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. FEBS Lett. **225**:163–167.
23. **Lever, M.** 1972. Colorimetric and fluorometric carbohydrate determination with *p*-hydroxybenzoic acid hydrazide. Biochem. Med. **7**:274–281.
24. **Love, D. R., R. Fisher, and P. L. Bergquist.** 1988. Sequence structure and expression of a cloned β-glucosidase gene from an extreme thermophile. Mol. Gen. Genet. **213**:84–92.
25. **Love, D. R., and M. B. Streiff.** 1987. Molecular cloning of a β-glucosidase from an extremely thermophilic anaerobe in *E. coli* and *B. subtilis*. Bio/Technology **5**:384–387.
26. **Lowry, O. H., N. J. Rosenbrough, A. L. Farr, and R. J. Randell.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
27. **Lüthi, E., D. R. Love, J. McAnulty, C. Wallace, P. A. Caughey, D. J. Saul, and P. L. Bergquist.** 1990. Cloning, sequence analysis and expression of xylan-degrading enzymes from the thermophile "*Caldocellum saccharolyticum*". Appl. Environ. Microbiol. **56**:1017–1024.
28. **MacKay, R. M., A. Lo, G. Willide, M. Zuker, S. Baird, M. Dove, F. Moranelli, and V. Seligy.** 1986. Structure of a *Bacillus subtilis* endo-β-1,4-glucanase gene. Nucleic Acids Res. **14**:9159–9170.
29. **Miller, R. C., N. R. Gilkes, N. M. Greenberg, D. G. Kilburn, M. L. Langsford, and R. A. J. Warren.** 1987. *Cellulomonas fimi* cellulases and their genes, p. 235–248. *In* J.-P. Aubert, P. Béguin, and J. Millet (ed.), Biochemistry and genetics of cellulose degradation. Academic Press, London.
30. **Millet, J., D. Petre, P. Béguin, O. Raynaud, and J.-P. Aubert.** 1985. Cloning of ten distinct DNA fragments of *Clostridium thermocellum* coding for cellulases. FEMS Microbiol. Lett. **29**:145–149.
31. **Montencourt, B. S.** 1983. *Trichoderma reesei* cellulases. Trends Biotechnol. **1**:156–161.
32. **Norrander, J., T. Kemple, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **10**:137–145.
33. **O'Neill, G. P., S. H. Goh, R. A. J. Warren, D. G. Kilburn, and R. C. Miller.** 1986. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene **44**:325–330.
34. **Rackwitz, H.-R., G. Zehertner, A.-M. Frischauf, and H. Lehrach.** 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene **30**:195–200.
35. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
36. **Saul, D. J., A. J. Spiers, J. McAnulty, M. G. Gibbs, P. L. Bergquist, and D. F. Hill.** 1989. Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids. J. Bacteriol. **171**:2697–2707.
37. **Saul, D. J., L. C. Williams, D. R. Love, L. W. Chamley, and P. L. Bergquist.** 1989. Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity. Nucleic Acids Res. **17**:439.
38. **Schauder, B., H. Blöckner, R. Frank, and J. E. G. McCarthy.** 1987. Inducible expression vectors incorporating the *Escherichia coli atpE* translational initiation region. Gene **52**:279–283.
39. **Schwarz, W. H., F. Grabnitz, and W. L. Staudenbauer.** 1986. Properties of a *Clostridium thermocellum* endoglucanase produced in *E. coli*. Appl. Environ. Microbiol. **51**:1293–1299.
40. **Shoemaker, S., V. Schweickart, M. Ladner, D. Gelfand, S. Kwok, K. Myambo, and M. Innis.** 1983. Molecular cloning of exocellobiohydrolase I derived from *Trichoderma reesei* strain L27. Bio/Technology **1**:691–696.
41. **Short, J. M., J. M. Fernandez, J. A. Sarge, and W. D. Huse.** 1988. λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. Nucleic Acids Res. **16**:7583–7600.
42. **Teather, R. M., and P. K. Wood.** 1982. Use of Congo Red-polysaccharide interactions in enumeration and characterization

of cellulolytic bacteria from bovine rumen. Appl. Environ. Microbiol. **43**:777–780.

43. **Van Tilbeurgh, H., M. Claeyssens, and C. K. De Bruyne.** 1982. The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of cellulolytic enzymes. FEBS Letts. **149**:152–156.
44. **Warren, R. A. J., B. Gerhard, N. L. Gilkes, J. B. Owolabi, D. G. Kilburn, and R. C. Miller.** 1987. A bifunctional exoglucanase-endoglucanase fusion protein. Gene **61**:421–427.
45. **West, A. W., A. Elzanowski, L.-S. Yeh, and W. C. Barker.** 1989. Homologues of catalytic domains of *Cellulomonas* glucanases found in fungal and *Bacillus* glycosidases. FEMS Microbiol. Lett. **59**:167–172.
46. **Wong, W. K. R., B. Gerhard, Z. M. Gou, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. Gene **44**:315–324.